ERIC Document Resume	See Instruction Sheet	DATE OF RESUME 8/15/65	DATE RECEIVED

A. PUBLICATION DATA

1. ED NUMBER	2. ORIGINATOR CONTROL NO(S).	3. IDENTIFICATION NO.	DO NOT WRITE IN SPACE BELOW
ED000537	7288		

4. AUTHOR(S):
Nelson, Wallace B.

5. TITLE OF DOCUMENT ABSTRACTED:
An Experiment with Class Size in the Teaching of Elementary Economics

6. PUBLICATION TITLE:
The Educational Record

7. EDITOR(S):
N.A.

9. PUBLISHER OR SOURCE
Kansas State University

8. PLACE OF ORIGIN
ORG. Kansas State
ST. University
CITY, STATE Manhattan, Kansas
TELE. EXT.
INST/LOC CODE

10. VOL.	11. PAGES	12. MO./YR.	13. TOTAL PP.
40	330-341	Oct. 59	12

MICROFILM COPY AVAILABLE YES ■ NO □

COPYRIGHT GRANTED YES ■ NO □

COPYRIGHT RELEASE FILED YES ■ NO □

TYPE OF REPORT
PROPOSAL □
IN-PROGRESS □
FINAL ■
FOLLOW-UP □

OTHER ERIC REFERENCES:

ERIC SATELLITE CODE

B. INDEXING DATA

14. SUBJECT AREA

15. TYPE OF ENDEAVOR
experimental study

16. INDEXING TERMS (6-12)

	teaching methods	examinations		
		students		
experimental studies	ecomics	achievement		
class size	effectiveness			

17. ABSTRACT (MAX. 250 WORDS)

Reviews experiments on teaching effectiveness as related to class size, and describes one conducted at Kansas State University from the fall semester 1956-57 and through the spring semester 1957-58. Four instructors were involved, each teaching both large and small sections in Economics I. Presents conclusions and qualifications. Stresses the need to define, appraise, and test intangibles which seem to represent vital educational objectives but which may be neglected, in whole or in part, by measuring student achievement in terms of objective-type examinations.

ERIC No.

18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION	20. DIVISION OR DEPT.	21. CONTRACT NO. / PROJECT NO.
	22. TOTAL COST OF PROJECT	
	23. DURATION OR PROJECT	U.S. Department of Health, Education, and Welfare Office of Education

ERIC Form 1

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Reprinted from the October 1959 issue of THE EDUCATIONAL RECORD published by the American Council on Education, Washington, D.C.

M-330-41 Vol. 40, No. 4

IV-21

An Experiment with Class Size in the Teaching of Elementary Economics[1]

WALLACE B. NELSON

AFTER REACHING a postwar low of 786 in the 1952–53 academic year, enrollment in the Principles of Economics course at Kansas State University increased 16 percent in each of the years 1953–54 through 1956–57 until, for the latter year, it stood at 1,461. Over this four-year period, enrollment increased 86 percent. To have absorbed this increased enrollment without additional staff would have increased the average class size from 25 to 46. Actually, teaching resources were augmented to the extent that the average class size in the fall semester of 1956–57 was held to 39.[2]

The sheer magnitude of the problem, and the rapidity with which it was obviously growing, dictated the necessity for casting about for ways and means of meeting it. Other schools were, of course, faced with the same problem. The University of Colorado began experimenting with a "mentor-mentee" system,[3] and the University of Nebraska undertook an experiment involving "specialization of instruction."[4] Both these experiments involved increasing class size, but neither apparently regarded class size per se as a relevent variable. Dr. Schmidt reports that the Nebraska experiment "was not designed to test the relative merits of the large versus the small section . . . ,"[5] and the staff of the Department of Economics at Colorado observe in their application for a research grant that "the small section is the superior form of organization for teaching."[6]

It is probable that any survey of faculty opinion would find a large majority sympathetic to the Colorado view. The difficulty is that there is, so far as the writer is able to determine, virtually no evidence to support this viewpoint, and considerable which clearly disputes it.

Strangely enough, the teaching of economics, while the primary vocation of people dedicated to scientific experimentation in the social sciences, seems not to have been subjected to much scientific investigation itself. There is no reference to the problem of increasing class size and its impact upon teaching effectiveness in the report of the American Economic Association on

[1] The author is indebted to Dr. George Montgomery, head of the Department of Economics and Sociology, Kansas State University of Agriculture and Applied Science. Without his generous support, in obtaining financial support and otherwise, this study could not have been conducted.

[2] The average in this case is misleading: it includes four experimental sections, two of more than 140 each and two of 20 or less, as well as other sections.

[3] "Program for Utilization of College Teaching Resources," an application for a research grant from the Fund for the Advancement of Education by the Department of Economics, University of Colorado (Mimeographed; n.d.).

[4] "Research in Teaching the Principles of Economics, Final Report," prepared by E. B. Schmidt, professor and chairman, Department of Economics, University of Nebraska (Mimeographed; n.d.).

[5] *Ibid.*, p. 3.

[6] "Program for Utilization of College Teaching Resources," p. 4.

7288

ED000537

"The Teaching of Undergraduate Economics." [7]

Professors Birney and McKeachie, commenting on this phenomenon as it relates to psychology, speculate that:

> It is probable that one reason for such a lack of research is that until recent times psychology classes were sufficiently small to allow the instructor the feeling of intimacy that is so necessary for the intuitive feeling of success as a teacher.[8]

In the early stages of the writer's interest in this problem, when it seemed logical to assume the likelihood of other experiments, letters were written to the departments of economics in seventeen colleges and universities whose enrollments would seem to pose the same problems as were beginning to face Kansas State.[9] Of the twelve which replied, none had conducted any such studies and none knew of others who had. Of several organizations interested in economic education, only one was aware of any such studies. The Joint Council on Economic Education suggested that the University of Connecticut had been experimenting with class size. An inquiry resulted in a letter which read, in part:

> For the past few years we have experimented on having six large sections of about 180 students each and one small section of about 40 students. I have taught both a large and a small section simultaneously and the *results have been pretty much as expected; the smaller section, because of the opportunity to ask questions immediately and discuss the answers to these questions, has absorbed more of the course and has achieved better performance.*[10]

A follow-up letter requesting information on course objectives, research design, and statistical techniques elicited from Lurie the delightfully candid observation that:

> I am afraid that either Dr. Allen or myself overstated the "scientific" nature of our class size experiment. It was little more than feeling that prompted me to suggest that the performance was substantially better in the smaller than in the larger sections.[11]

E. E. Hale, chairman of the Department of Economics at the University of Texas, after indicating that "we have conducted no such experiment" (comparing effectiveness of large versus small classes), went on to observe that "our experience has been, over the years, that teaching in the elementary course can be done *far more effectively with small size sections.*" [12]

The only experiment designed to relate class size to teaching effectiveness involving economics which has come to the writer's attention is that reported by

[7] *American Economic Review*, December 1950, Supplement, entire issue.

[8] R. Birney and W. McKeachie, "The Teaching of Psychology: A Survey of Research Since 1942," *Psychological Bulletin*, January 1955, p. 51.

[9] The seventeen colleges contacted were Brooklyn College, University of Chicago, University of Colorado, University of Illinois, State University of Iowa, Iowa State University of Science and Technology, University of Kansas, University of Michigan, Michigan State University, University of Minnesota, University of Nebraska, City College of New York, New York University, Oklahoma State University of Agriculture and Applied Science, University of Oklahoma, University of Texas, and University of Wisconsin.

[10] Letter to the writer from Melvin Lurie, instructor of economics, University of Connecticut, July 3, 1956; italics added.

[11] This reference (and others in like vein) are not offered as criticism. They illustrate a point which seems most important: that our affinity for small-sized sections is not supported by concrete evidence. If small classes are to be maintained, especially in introductory courses, administrators and legislators are entitled to—and will likely demand—more objective evidence of their superiority than it now seems possible to supply.

[12] Letter from Dr. Hale to the writer, June 12, 1956, italics added.

Macomber and Siegel.[13] Here the comparison was between a large television class (141 students) and "small" control sections of 48 students each. The course was Introductory Economics. No significant difference in student achievement resulted.

While virtually nothing appears to have been done in the way of experimentation with class size and teaching effectiveness in economics, there is a considerable literature dealing with the problem of class size generally. Beginning with the pioneer studies by Rice[14] in 1895, some 260 studies were recorded by 1940.[15] Of this number, perhaps 70 can lay some claim to being "scientific experimentation."[16] Of this 70, the great majority represent experiments conducted at the primary or secondary level. The first true scientific experiment seems to have been conducted by Breed and McCarthy[17] in 1916; prior experiments were handicapped by the lack of intelligence tests and evaluative techniques.

Class size studies at the college level began in 1923, with the work of Mueller[18] and Edmondson and Mulder.[19] Mueller found significant differences in achievement between "large" classes of 40 students and "small" classes of 20 students in introductory psychology. The difference favored the small classes. Edmondson and Mulder, comparing sections of 45 and 109, found no significant differences in student achievement. The course was in education—An Introduction to High-School Problems. The best and certainly the most comprehensive study of this period is that of Hudelson at Minnesota.[20] Summarizing the research (including his own) done up to that time at all instructional levels, Hudelson concludes that "In the light of all available evidence, class size seems to be a relatively minor factor in educational efficiency, measured in terms of student achievement."[21]

Other studies conducted subsequent to this date but prior to 1932 and reported by Brown reinforce this conclusion.[22] The decade 1920 to 1930 seems to have been the period of most active interest in the problem of class size, judging from the frequency of publication.[23] After 1930 interest seems to have languished. Holland observed in 1954 that "this important area [of class size research], has been all but abandoned

[13] F. G. Macomber and Lawrence Siegel, "A Study in Large-Group Teaching Procedures," *Educational Record*, July 1957, pp. 220–29.

[14] J. M. Rice, "Educational Research: A Test in Arithmetic," *The Forum*, October 1902, pp. 281–97.

[15] Fred von Borgersrode, "Class Size," *Encyclopedia of Educational Research*, ed. Walter S. Monroe (New York: Macmillan Co., 1941), p. 197.

[16] Olis G. Jamison, "Class Size," *Encyclopedia of Modern Education*, ed. Harry N. Rivlin (New York: Philosophical Library, 1943), pp. 143–45.

[17] S. F. Breed and Grace D. McCarthy, "Size of Class and Efficiency of Teaching," *School and Society*, December 1916, pp. 965–71.

[18] A. D. Mueller, "Size of Class as a Factor in Normal School Instruction," *Education*, December 1924, pp. 203–7.

[19] J. B. Edmondson and J. F. Mulder, "Size of Class as a Factor in University Instruction," *Journal of Educational Research*, January 1924, pp. 1–12.

[20] Earl Hudelson, *Class Size at the College Level* (Minneapolis, Minn.: University of Minnesota Press, 1928).

[21] *Ibid.*, p. 270.

[22] A. E. Brown, *The Effectiveness of Large Classes at the College Level: An Experimental Study Involving the Size Variable and the Size-Procedure Variable* (State University of Iowa Studies in Education, Vol. VII, No. 3 [Iowa City: State University of Iowa, 1932]). For a summary of the literature at the elementary and secondary school level to 1954, see Howard E. Blake, "Class Size: A Summary of Selected Studies in Elementary and Public Schools" (Unpublished Ed.D. dissertation, Teachers College, Columbia University, 1954), 119 pp.

[23] Von Borgersrode, *loc. cit.*

for the last 15 years."[24] Renewed interest is evident by the late 1940's, probably reflecting (at the college level) the bulge in enrollments occasioned by the GI bill.

The most elaborate study of this period seems to have been that reported by Macomber and Siegel. Supported by a four-year, $285,600 grant from the Fund for the Advancement of Education, "the primary purposes of the study . . . were to . . . determine the effectiveness of large-group instructional procedures, including the use of television, for course instruction, and to find ways and means of improving upon present large-group practices in college teaching."[25]

Courses involved in the experiment included Foundations of Human Behavior, Physiology, Introduction to Economics, Air Science, Introductory Sociology, Human Biology, Introduction to Business, Geology, Business and Government, Composition and Literature, Geography, Elementary Psychology, French (Reading Comprehension), Chemistry, Algebra, Zoology, History of Civilization, Physics, Classics, and Government.[26] "The differences in achievement of students in the experimental and control sections are statistically significant in only two of the courses for which data were available—biology and geology."[27] In the case of biology, there was no control section. The TV section was compared "with a large lecture section taught during the previous semester. . ."[28]

A study at the University of Oklahoma involved large classes of 309 and 332 compared with small classes of 23 and 31, respectively. The course was American Government. "The differences in the achievement of the students in the large and small sections . . . were . . not statistically significant."[29]

An experiment with a beginning geography course is reported by Perry.[30] Here large sections of approximately 125 students were compared with small sections of not more than 30 students. "Statistically, there was no significant difference between the examination results of the students in the large lecture groups and the students in the small control groups."[31]

While each of the experiments summarized above may be open to criticism in certain particulars, they are all of uniformly high quality generally, and the overwhelming weight of the findings is to the effect that large classes of from 40 to 250 are as effective as small classes of less than 40, where effectiveness is defined as student achievement on objective-type examinations. Still, this conclusion goes down hard. A faculty committee at Ball State Teachers College, after observing Hudelson's findings that "it makes no difference what the class size is, at least from say 20 to 150 or even more," and reflecting that he probably was not "very popular with his colleagues for awhile" but that "facts are facts and he seems to have a lot of facts," ends up by recommending that maximum class

[24] John B. Holland, "The Image of the Instructor as It Is Related to Class Size," *Journal of Experimental Education*, December 1954, pp. 171–77. This is one of the few studies which have attempted to relate class size to a variable other than "student achievement as measured by objective type examinations." It represents a virgin area of potential research. In this connection see also Macomber and Siegel, *op. cit.*

[25] Macomber and Siegel, *op. cit.*, p. 221.

[26] *Ibid.*, p. 224.

[27] *Ibid.*, p. 223.

[28] *Loc. cit.*

[29] J. H. Rohrer, "Large and Small Sections in College Classes," *Journal of Higher Education*, May 1957, pp. 275–80.

[30] Robert F. Perry, "A Teaching Experiment in Geography," *Journal of Geography*, March 1957, pp. 133–35.

[31] *Ibid.*, p. 134.

size at Ball State Teachers College "be held to 30."[32]

Can elementary economics be taught as effectively in large classes as in small? Aside from the inferences which may be drawn from the experiments involving the teaching of other disciplines, no answer seemed to be available. The present experiment was designed to provide at least a tentative one.

The experiment

The ideal research design of an experiment to test the hypothesis that teaching effectiveness is not importantly related to class size is deceptively simple. One need only define objectives, conduct large and small classes simultaneously—with all relevant variables save class size (and the associated teaching method) held constant or controlled—test for achievement, and compare the results. If, under these circumstances, the small classes attain significantly higher scores, then the staff of the University of Colorado must be correct: the small section is the superior form of organization for teaching.[33]

The problem of adequately defining objectives is perhaps the most difficult, and certainly one of the most essential, requirements of such an experiment. No claim is made that the objectives set down for this experiment are adequate as guides for the Principles course. Quite the contrary: their deficiencies in this respect are obvious and glaring. There was, however, a fairly precise set of objectives, and a systematic attempt to test for their achievement.[34]

The research design, the course objectives, and the pre- and post-examination were all prepared in the summer of 1956. The experiment began in the fall semester of 1956–57, and continued through the spring semester of 1957–58. Four different instructors were involved, each teaching both a large and a small section of Economics I: the first half of the two-semester-sequence Principles course. All experimental sections were Monday, Wednesday, and Friday morning hours. To ensure the same caliber of student group, each instructor's large and small sections were matched, or made identical, insofar as possible, in three respects. These were the student's school (to avoid, for example, having predominantly business administration majors in one section and engineers in the other); student classification (freshman, sophomore, junior, senior); and sex. Mechanically, this matching was accomplished by selecting from a much larger group the 18 to 20 students who would comprise the small section, making the small section as nearly identical to the large as possible. The degree to which this process was successful for the seven parts of the experiment is indicated in Table 1.

While, as the table shows, this matching process was not perfect, a substantial degree of similarity between large and small sections was achieved. Aside from selectivity factors operating within these categories, the large and small sections were as nearly alike in terms of student composition as they could be made. Any remaining differences in student ability between sections were eliminated statis-

[32] "A Study of Class Size Control at Ball State Teachers College," *Teachers College Journal*, December 1950, pp. 47–49.

[33] One such case would not be sufficient. The writer is of the opinion that the seven cases which constitute the present experiment provide prima facie evidence on which the conclusions reached may rest until they are disputed by more objective rebuttal than now seems available.

[34] These objectives, and the examinations designed to test for their achievement, are available to anyone interested.

TABLE 1: SCHOOL, CLASSIFICATION, AND SEX DISTRIBUTION, IN PERCENTAGES, OF STUDENTS COMPRISING EXPERIMENTAL SECTIONS

EXPERIMENTAL PERIOD	INSTRUCTOR	SECTION SIZE	SCHOOL				CLASSIFICATION						SEX	
			Eng.	A & S	Ag.	HE	Fr.	So.	Jr.	Sr.	Grad.	Spec.	M	F
Fall 1956–57	A	20	40	50	10	0	15	75	10	0	0	0	90	10
		138	42	34	24	0	9	78	13	0	0	0	93	7
	B	16	44	31	25	0	18	63	18	0	0	0	90	10
		141	42	33	25	0	11	73	16	0	0	0	94	6
Spring 1956–57	A	20	45	35	20	0	50	45	5	0	0	0	90	10
		94	49	31	17	3	49	44	4	2	0	1	90	10
	B	20	25	40	25	10	25	50	20	5	0	0	85	15
		90	24	42	28	5	38	42	18	2	0	0	89	11
Fall 1957–58	A	17	53	35	12	0	5	77	18	0	0	0	100	0
		109	55	32	13	0	9	74	15	1	0	0	100	0
	C	17	53	42	5	0	0	95	5	0	0	0	100	0
		94	53	34	13	0	11	70	16	3	0	0	100	0
Spring 1957–58	D	19	37	42	16	5	37	53	10	0	0	0	79	21
		85	41	39	18	2	37	50	8	2	1	2	81	19

tically through analysis of covariance techniques.[35]

To the extent that the process described encompassed the relevant variables, the requirements of a valid experiment would seem to have been met. The classes were all scheduled for popular Monday, Wednesday, and Friday morning hours, so that the selection of students by scheduling some sections at unpopular hours was minimized. The instructor variable was at least partially controlled by having the same instructor teach both a large and a small section.[36] The student variable was controlled by matching large and small sections and by statistical techniques designed to

[35] The assumption here was that student ability was adequately measured by cumulative grade-point average. Scores on the American Council on Education Psychological Examination were considered less useful for this purpose than cumulative grade-point average, because ACE scores neglect application. A student with a relatively lower ACE score may perform better than one with a higher ACE because he tries harder.

[36] Reindigs Fels, in a letter to the author, observed in this connection that the outcome of any experiment with class size was probably a function chiefly of the instructor. To quote Fels: ". . . the outcome of your experiment will depend on the man who teaches the large section. If he is outstanding, the results will be better than with small sections, but if he is average, the results will be worse."

Taken literally this would seem to imply that an outstanding teacher could always be expected to perform better with large classes than with small. I see no a priori reason why this should be true.

If, however, it means that teachers who are effective before large groups and ineffective before small groups should teach only large groups (and vice versa) this is scarcely debatable, but neither is it particularly relevant unless most instructors are competent in only one situation or the other. I suspect that most average, reasonably competent instructors (and most outstanding ones) are, or can be, as effective in one situation as in the other. In the present experiment, four instructors demonstrated that they were as effective (or ineffective) in classes of 85 to 140 as in classes of 16 to 20.

TABLE 2: IMPROVEMENT SCORES

EXPERIMENTAL PERIOD	INSTRUCTOR	ADJUSTED MEANS*		DIFFERENCE OF MEANS	COMPUTED† t	APPROXIMATE PROBABILITIES‡
		Small	Large			
Fall 1956–57	A	6.70	7.37	−0.67	+0.772	.44
	B	9.79	8.57	1.22	1.275	.20
Spring 1956–57	A	6.90	6.98	−0.08	+0.089	>.50
	B	6.64	7.25	−0.61	+0.729	.46
Fall 1957–58	A	8.11	7.91	0.20	0.199	>.50
	C	7.84	8.18	−0.34	+0.339	>.50
Spring 1957–58	D	7.19	6.87	0.32	.292	>.50

*These means have been adjusted for variance in grade-point averages which were used as a covariate.
†See George W. Snedecor, *Statistical Methods* (5th ed.; Ames, Iowa: Iowa State College Press, 1956), pp. 400–403.
‡Probability of obtaining a t which exceeds the computed t in a two-tailed test under the null hypothesis.

eliminate such differences as this matching process neglected.

The small section, potentially at least, was a discussion-recitation-group situation. The large group was more nearly straight lecture. The same material was covered in both classes, and both classes were given the same true-false and/or multiple-choice examinations. Two measures of student achievement were employed. The first was called the "improvement score." An examination was given the first day of the class and again at the end of the semester. If a student answered 20 questions correctly to begin with, and 30 at the end, his improvement score was 10. A second, and perhaps more accurate, measure of student achievement was the sum of the raw scores in four major examinations, the last of which included the post-test.[37]

The mean improvement scores for the seven parts of the experiment, adjusted for variation in student ability (as measured by grade-point averages), are given in Table 2.

For this achievement measure, there is clearly no difference between large and small sections.[38]

Using final grades as a measure of student achievement, the picture is not quite so clear cut, but the weight of the evidence is still heavily in favor of the proposition that section size is not an important determinant of student achievement. Final grades (mean of the sum of the raw scores in four major exams) for the seven parts of the experiment, adjusted for variation in student ability, are given in Table 3.

For Instructor A, the difference of means between large and small sections is never as great as one percentage point, and two times out of three the difference is in favor of the large section. For Instructor B, however, the difference approaches four percentage points in the fall semester, and favors the small sec-

[37] In statistical terminology, the hypothesis being tested can be stated as follows: For a given instructor and a given semester, the mean of the hypothetical population of which students in the small section are members is equal to the mean of the hypothetical population of which students in the large section are members.

[38] In statistical terminology: For this achievement measure, the probabilities of obtaining the t values are so large that the differences in averages from one section size to another might well be attributed to experimental error. Hence there is insufficient evidence to reject the hypothesis of equal population means.

TABLE 3: FINAL GRADES

EXPERIMENTAL PERIOD	INSTRUCTOR	ADJUSTED MEANS*		DIFFERENCE OF MEANS	COMPUTED† *t*	APPROXIMATE PROBABILITIES‡
		Small	Large			
Fall 1956–57	A	67.92	68.77	−0.85	+0.517	>.50
	B	73.81	70.11	3.70	2.031	.04
Spring 1956–57	A	73.01	72.31	0.70	0.405	>.50
	B	77.22	75.31	1.91	1.096	.28
Fall 1957–58	A	73.80	73.92	−0.12	+0.075	>.50
	C	72.86	75.91	−3.05	+1.837	.07
Spring 1957–58	D	71.51	70.56	0.95	0.491	>.50

*These means have been adjusted for variance in grade-point averages which were used as a covariate.
†See George W. Snedecor, *Statistical Methods* (5th ed.; Ames, Iowa: Iowa State College Press, 1956), pp. 400–403.
‡Probability that *t* exceeds the absolute value of the computed *t* in a two-tailed test under the nul. hypothesis.

tion both times. The evidence for rejecting the hypothesis of equal means is fairly strong (Probability=.04) for the fall 1956–57 period, but for the spring 1956–57 period the difference in sample means might be attributed to experimental error. Hence, the argument that the final grades for small sections will be higher than that for large sections, even with Instructor B, is not clear cut.

A peculiar phenomenon occurred for the small sections taught by Instructor B that did not occur for any of the other sections. The correlations between grade-point average and final grade were .20 and .02, respectively, for the fall and spring semesters for Instructor B's small sections. For *all* other sections (large and small) this correlation ranged between .50 and .72. This would point to the phenomenon of either the "poor" students performing above expectation or the "good" students performing below expectation, or both. The over-all effect may have been a slight increase in the mean for these small sections.

For Instructor C, there is some evidence (Probability=.07) that the large section did better than the small section. No explanation for this phenomenon was found, and at this probability level one is hesitant to declare that students in large sections taught by C will do better than those in small sections. For instructor D, no significant difference between large and small sections resulted.

By way of summary, then, five times out of seven, the difference in performance between large and small sections, as measured either by improvement scores or final grades, was so small that it might well be attributable to experimental error. In one case (Instructor B, fall semester, 1956–57), there seemed to be a significant difference in performance, as measured by final grades, in favor of the small section. This difference did not repeat, however, even for this instructor. In the remaining case (Instructor C, fall semester, 1957–58), the large class achieved a higher average final grade than the small, although not at a probability level that would allow one to predict with much confidence that this difference would re-occur.

It would seem fair to conclude that if class size affected student achievement in the present experiment, this ef-

fect was so small as to be negligible.* For the instructors involved, the experiment thus leads to a rejection of the hypothesis that the small section is the superior form of organization for teaching elementary economics. Student achievement was as great in large sections of from 85 to 140 as in small sections of 16 to 20.

Qualifications and conclusions

An unqualified conclusion to the effect that large classes are as effective as small could, if taken literally, have serious consequences. What follows is an attempt to state with more precision what the present experiment seems to prove, to qualify these conclusions to the extent that they seem to require qualification, and to make explicit what, in the writer's view, the experiment does not prove.

At the outset it should be observed that no unqualified assertion to the effect that large classes are as effective as small has been made, and none is warranted. If the experiment was valid, however, it would seem to warrant the following qualified conclusion: *For the four instructors involved, teaching the course they taught and using the teaching techniques and testing methods they employed, large classes of from 85 to 140 can be taught as effectively as small classes of 16 to 20.*

There would seem to be three major grounds on which the validity of this conclusion, or its significance, might be challenged. First, it might be argued that the experiment was faulty and, hence, the conclusion unwarranted. Second, the charge may be that the conclusion, if warranted, is of limited significance, applying as it does only to Kansas State, and here only for the instructors involved. Finally, and in my view most importantly, it may be objected that the experiment, even if valid and the conclusions widely applicable, is not very significant in that it concerns itself with the effect of class size on student achievement as measured by objective-type examinations, and thus ignores certain perhaps more intangible but equally important factors.

Within the first category of possible criticism—that the experiment was faulty—the most reasonable objection would seem to center on the variables, other than class size, which might have influenced the result. The assumption was that the effect of all relevant variables save class size (and the associated teaching method) was the same for each pair of large and small sections. If an important relevant variable was *not* accounted for, however, this would influence the result and prejudice the conclusion. Suppose, for example, that this relevant, but uncontrolled, variable operated to the advantage of the large section. Had it been controlled, and its influence eliminated, the small section would have been demonstrably superior and our prejudices confirmed.[10]

Were there such important uncontrolled variables? I think not. The same instructor, teaching comparable sections (in terms of student ability), using the same text, covering the same material, and evaluating the result with the same objective-type examinations, would seem to have eliminated or controlled all the important relevant variables.

The instructor variable was obviously

* In statistical terminology: If there was any differential effect on either improvement scores or final grades due to section size for the experiment as a whole, this effect was so small that these experiments could not detect such a difference at a probability level that would lead one to reject, with a high degree of confidence, the hypothesis of no difference in population means.

[10] This potentially cuts both ways of course. Uncontrolled variables operating in favor of small sections might have prejudiced the result. If so, had this (these) variable(s) been controlled, the large sections might have been demonstrably superior to the small.

controlled. There may be more question about the student variable. On the assumption that selectivity factors operate between schools,[41] an effort was made to make the large and small sections comparable in this respect. The same was true for student classification and sex. Are there other relevant variables beyond these affecting student ability? The University of Nebraska experimenters thought so and attempted to evaluate the effect (on student achievement) of outside employment, extracurricular activities, major interests in high school, type of student housing, family background, and reasons for taking the course.[42]

While there seems little reason to object to such exhaustive analysis, except perhaps as one is required to finance it, there is at the same time some question of its value. If cumulative grade-point average is a valid measure of student ability, and if classes can be standardized through selection of students and/or statistical manipulation, there would seem to be no requirement for assessing, for example, the student's family background. This factor is included in cumulative grade-point average, as are the others.[43] And even on the assumption that important uncontrolled variables were present and did influence the result, there would seem to be a prima facie case to the effect that they would continue to influence it in the same way, so that the conclusion is still valid: no significant difference between large and small classes.

A more telling objection in this category might be that the experiment did not really compare the effectiveness of small versus large classes. An examination of, and reply to, this charge can be made in terms which will also serve to assess the second most likely criticism: that the conclusion is not very significant, even if true, because it applies only to Kansas State, and here only to the instructors involved.

Presumably the advantages of small classes stem primarily from the intimate student-instructor relationship which they promote, or at least permit. The ideal is perhaps typified by Garfield's conception of a university as "Mark Hopkins on one end of a log and a student on the other," or the European tutorial system.[44]

Did the small classes which were a part of the experiment approach this ideal? Probably not. As was stated previously, the small section, potentially at least, was a discussion-group situation. In fact, however, the potentialities of the discussion-group technique were probably not fully exploited. The same is undoubtedly true of the large lecture group. If the charge, then, is that the conclusion is not valid because full advantage was not taken of the opportunities made available by the small classes, the answer is clear. It must be conceded that the experiment did not compare the perfect small-group teaching situation with the perfect large-group teaching situation. This would be an in-

[41] The two major groups which together comprise a sizable majority of total Principles of Economics enrollment at Kansas State are engineers and business administration majors. Undergraduates studying business and commerce have, on the average, a lower intelligence level than any other group of predominantly male students, with the exception of those majoring in physical education. Engineers are among the highest. See Dael Wolfle, *America's Resources of Specialized Talent* (New York: Harper & Bros., 1954), chap. 7.

[42] "Research in Teaching the Principles of Economics: Final Report," prepared by Dr. E. B. Schmidt, professor and chairman of the Department of Economics, University of Nebraska (n.d.), pp. 4–10.

[43] For some reason "it was not possible to assign students to the experimental and non-experimental sections upon a scientific basis" in the Nebraska experiment. *Ibid.*, pp. 2, 10.

[44] William S. Rouverol, "The Tutorial System," *AAUP Bulletin*, Winter 1954–55, pp. 593–605.

teresting experiment. But if all the professors of *academia* and of education were to call off their feud, and each devote his time to absorbing everything the other had to offer, it is doubtful whether any would even then be qualified to conduct it.

Both the large and the small classes which were a part of the experiment are probably fairly typical of the large and small classes in Principles of Economics at hundreds of other colleges and universities throughout the United States. I believe this to be the case. If so, the criticism that this was not really a valid comparison of large and small classes (in the sense noted above) while true, is irrelevant. Further, the fact that, scientifically, the conclusion applies only to Kansas State, and here only to the instructors involved, may be misleading. If conditions elsewhere are similar to those at Kansas State, as I think, the conclusion is widely applicable. The average, reasonably competent instructor in economics can teach Principles of Economics as effectively to classes of 85 to 140 as to classes of 16 to 20. It is at this point that the third objection may be posed. "Effectiveness" has been defined as student achievement as measured by objective-type examinations. If student achievement as measured by objective-type examinations is an adequate measure of all that we can reasonably hope to achieve by means of the educational process, then we should proceed posthaste (in introductory economics) to large classes. But it may not be.

Wall, using Trytten's findings on the undergraduate origin of physics Ph.D.'s,[45] suggests that "the general physics lectures may be better in the large university than in the small college. The students may learn more physics. They may do better on exam[illegible]ions, and the percentage of outstanding students may be larger. But when they are through with the course, they do not choose to go on with physics. . ."[46] Developing a possible explanation for this phenomenon, he continues:

> With 20 or 30 students in the class, the relationship between the teacher and his students may be personal and vital; with 250 students in the class it must of necessity become impersonal. . . . The student becomes a mere spectator instead of an active participant. This difference in point of view is fundamental, *but I doubt that it can be detected by the ordinary tests and examinations.* If the student is to get the meaning and value out of a course in physics, or any other course for that matter, he must immerse himself in the stream, not stand watching on the bank. But this latter position appears to be precisely that taken by the great majority of students in a large lecture section.[47]

The student's value system, his ability to think critically and to communicate effectively, and his tolerance of convictions which do not square with his own, all would seem to be among the intangibles which represent vital educational objectives but which may be neglected, wholly or in part, by the crude measuring device of objective-type examinations. These intangibles need to be more carefully defined, their absolute and relative importance appraised, and means devised for testing the degree to which they are being achieved. Until this is done, any administrative decision to preserve small classes (or to go to large) will be in part indefensible because we have not yet systematically appraised the effectiveness of large versus small classes with effectiveness defined to include all

[45] M. H. Trytten, "The Undergraduate Origin of Physics Ph.D's," *American Journal of Physics*, 1947, pp. 330–33.

[46] C. N. Wall, "Concerning the Teaching of Physics," *American Journal of Physics*, 1949, p. 264.

[47] *Ibid.*, p. 265. Italics added.

(or most) of the obviously relevant educational objectives. Even so, the burden of proof would seem to have been decisively shifted to the proponents of small sections. Among other possible alternatives, a policy of teaching most of the elementary economics course in sections of up to 150 seems defensible. In addition, a number of small sections, reserved for exceptional students with the brains and ability, the drive, and the communication skills required to exploit the discussion potential of such sections, may be in order.